

# Predlog modela za predviđanje koncentracije suspendovanih (PM<sub>2.5</sub>) čestica u vazduhu

## Proposing the Predictive Model for Airborne PM<sub>2.5</sub> Concentration

Filip Nastić

Fakultet inženjerskih nauka Univerziteta u Kragujevcu

**Rezime** - Rastući broj istraživanja ukazuje na negativan uticaj suspendovanih (PM) čestica na zdravlje ljudi. Jedan od načina da se izbegnu njihove negativne posledice, jeste blagovremena predikcija koncentracije PM<sub>2.5</sub> u vazduhu. Znajući časovnu koncentraciju, građani bi mogli organizovati svoje dnevne aktivnosti sa ciljem smanjenja njihovog izlaganja intenzivnom zagađenju. U cilju formiranja optimalnog modela za časovnu predikciju koncentracije PM<sub>2.5</sub> u vazduhu, analizirane su prediktivne performanse tri različita algoritma mašinskog učenja: „Random forest“, „XGBoost“ i „Light gradient boosting machine“. Koristeći pomenute algoritme mašinskog učenja stvoreni su modeli koji koristeći meteorološke i hronološke podatke mogu izvršiti predikciju koncentracije PM<sub>2.5</sub> na časovnom nivou sa zadovoljavajućom tačnošću. Podaci o koncentraciji PM<sub>2.5</sub> su prikupljeni laserskim senzorom u gradu Kragujevcu, čija su očitavanja preuzeta sa sensor.community otvorene baze podataka. Evaluacija modela je izvršena koristeći koeficijent determinacije (R<sup>2</sup>), osrednjenu apsolutnu grešku (MAE) i koren srednje kvadratne greške (RMSE).

**Ključne reči** - Zagađenje vazduha, časovna predikcija, mašinsko učenje, PM<sub>2.5</sub>, zdravlje ljudi

**Abstract** - Increasing number of studies indicate the negative influence of Particulate Matter on human health. One of the ways to avoid their negative consequences is a timely prediction of airborne PM<sub>2.5</sub> concentrations. Knowing hourly PM<sub>2.5</sub> concentrations, people could organize their daily activities to reduce exposure to intensive pollution. With the goal to train an optimal predictive model, the predictive performances of three machine learning algorithms were analysed: „Random forest“, „XGBoost“, and „Light gradient boosting machine“. Using mentioned regression algorithms in combination with meteorological and chronological data, the models were trained to predict hourly airborne PM<sub>2.5</sub> concentrations with relatively high accuracy. The data about airborne PM<sub>2.5</sub> concentrations were collected using the laser sensor in the city of Kragujevac, Serbia. The trained models were evaluated using the coefficient of determination (R<sup>2</sup>), mean absolute error (MAE), and root-mean-square error (RMSE).

**Index terms** - Air pollution, hourly prediction, human health, machine learning, PM<sub>2.5</sub>

### I UVOD

Zagađenje vazduha je postao jedan od najvećih problema današnjice. Aktuelna istraživanja ukazuju na štetan uticaj zagađenja na životnu sredinu, zdravlje ljudi i kvalitet života [1]. Prema Svetskoj zdravstvenoj organizaciji, zagađenje je odgovorno za oko 6.7 miliona prevremenih smrti godišnje [2], pri čemu je oko 4 miliona prevremenih smrti godišnje izazvano zagađenjem PM<sub>2.5</sub> [3]. Suspendovane (PM) čestice predstavljaju sve čvrste i tečne čestice koje se mogu naći u vazduhu. Razlikuju se:

- PM<sub>10</sub> - čestice u vazduhu prečnika manjeg od 10 µm,
- PM<sub>2.5</sub> - čestice u vazduhu prečnika manjeg od 2.5 µm.

PM<sub>2.5</sub> predstavljaju veliku pretnju po ljudsko zdravlje, jer one mogu, zahvaljujući svojim malim dimenzijama, preko pluća da dospeju u krvotok [4]. Problemi sa zdravljem koje izaziva izloženost PM<sub>2.5</sub> su bolesti respiratornog i kardiovaskularnog sistema i kancer pluća [5].

Trenutno je poznato da se u Srbiji oko 70% električne energije dobija iz termoeletrana na uglj [6]. Ako se uzme u obzir uticaj uglja na životnu sredinu i kvalitet vazduha, starost postojećih termoeletrana, zaključuje se da su građani Republike Srbije izloženi značajnoj količini zagađenja. Ideja ovog rada jeste da se razvije model za predviđanje časovne koncentracije PM<sub>2.5</sub> u vazduhu. Blagovremena predikcija zagađenja može ponuditi građanima mogućnost pravilne organizacije svojih obaveza i aktivnosti sa ciljem smanjenja izloženosti visokim koncentracijama ovih čestica.

Rad [7] je analizirao uticaj meteoroloških faktora na koncentraciju PM<sub>2.5</sub> u vazduhu širom Kine. Navedeno je da spoljna temperatura, brzina vetra, relativna vlažnost, precipitacija i atmosferski pritisak imaju značajan uticaj na koncentraciju ovih čestica. Istraživanje [8] je potvrdilo pomenute činjenice baveći se uticajem temperature, brzine vetra, apsolutne vlažnosti vazduha i precipitacije na koncentraciju PM<sub>2.5</sub> tokom letnje, zimske i jesenje sezone. Postoje različiti indikatori vlažnosti vazduha, dok je rad [9] zaključio da se relativna vlažnost može koristiti kao dobar indikator koji je u jakoj korelaciji sa koncentracijom PM<sub>2.5</sub> u vazduhu. Baveći se zavisnošću koncentracije PM<sub>2.5</sub> u vazduhu i spoljašnje temperature, rad [10] je došao do zanimljivih rezultata. Pokazano je da je spoljašnja temperatura u pozitivnoj korelaciji

sa koncentracijom predmetnih čestica, iz razloga što visoke temperature pozitivno utiču na hemijski proces formiranja PM u vazduhu. Istraživanje [11] daje kontradiktorne zaključke, ukazujući na negativnu korelaciju između temperature i koncentracije suspendovanih čestica izazvanu intezivnijim sagorevanjem goriva za potrebe grejanja usled nižih temperatura. Može se zaključiti da znak korelacije u slučaju temperature zavisi od toga da li se ona posmatra tokom ili van grejne sezone, što ukazuje na još jedan uticajni parametar, faktor grejne sezone. Takođe, ljudska aktivnost (saobraćaj, poljoprivreda, industrija, itd.) imaju značajan uticaj na koncentraciju suspendovanih čestica [12]. Ljudski uticaj na koncentraciju suspendovanih čestica u vazduhu se može opisati faktorom "Sat" koji opisuje promenu inteziteta ljudskih aktivnosti tokom dana.

Poređenjem performansi različitih regresionih algoritama sa aspekta predviđanja koncentracije  $PM_{2,5}$  u vazduhu, rad [13] je pokazao da su Random forest i Gradient boosting regressor algoritmi dali najveće koeficijente determinacije koji iznose 0,9872 i 0,9983 respektivno. Sa istim ciljem, rad [14] je proveravao prediktivne performanse regresionih algoritama u kombinaciji sa AdaBoost algoritmom (klasifikacioni meta-algoritam). Ponovo, Random forest i XGBoost algoritmi su se pokazali kao najbolji, rezultirajući koeficijentom determinacije od 0,922, pri čemu je korišćenjem AdaBoost algoritma postignut skroman skok u performansama (0,3%). Rezultati dati u istraživanju [15] ukazuju na još jedan regresioni algoritam koji može pružiti dobre prediktivne performanse, a to je CatBoost algoritam. U ovom istraživanju poređene su performanse Arima-Artificial neural network, Arima-Support vector machine, Decision tree i CatBoost algoritama sa aspekta predikcije koncentracije  $PM_{2,5}$  u vazduhu, gde je CatBoost nadmašio ostatak rezultirajući koeficijentom determinacije od 0,95. Rad [16] je analizirao mogućnost predikcije koncentracije  $PM_{2,5}$  u vazduhu koristeći Light gradient boosting machine algoritam poredivši njegove evulacione pokazatelje sa rezultatima iz referentnih radova. Ovaj algoritam se pokazao kao najbolji, rezultirajući koeficijentom determinacije od 0,84 za dnevnu predikciju.

Na osnovu priložene literature, mogu se usvojiti uticajni

parametri i algoritmi koji će se koristiti u analizi:

- uticajni parametri koji će se koristiti u predmetnoj analizi u ovom radu su: Spoljašnja temperatura, brzina vetra, relativna vlažnost, atmosferski pritisak, faktor „Sat“, faktor grejne sezone i koncentracija  $PM_{2,5}$  24 h pre momenta predikcije;
- regresioni algoritmi koji će se koristiti u predmetnoj analizi u ovom radu su: Random forest, XGBoost, CatBoost i Light gradient boosting machine.

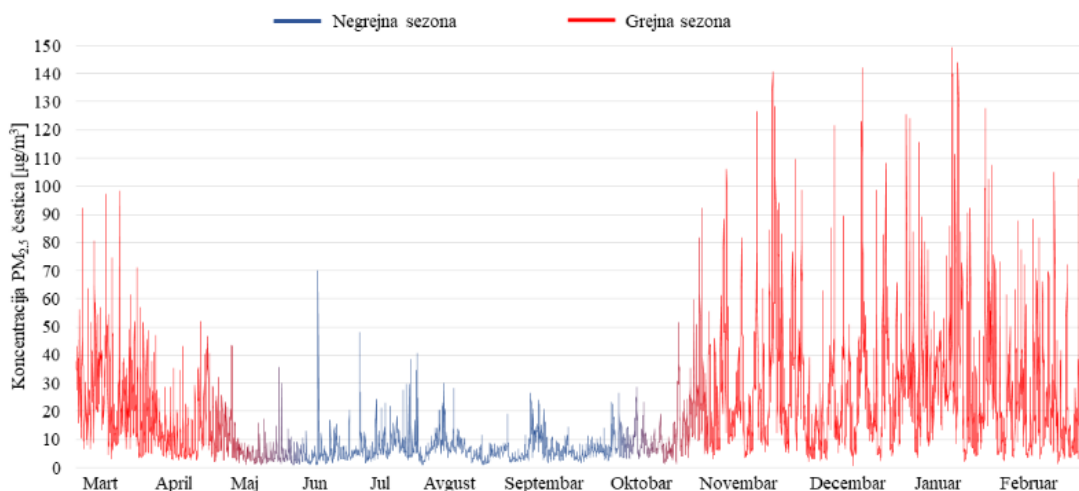
## II MATERIJAL I METODOLOGIJA

Ovaj rad je istraživao mogućnost predikcije koncentracije  $PM_{2,5}$  u vazduhu na primeru grada Kragujevca. Posmatrana lokacija je jedan od najvećih gradova u Republici Srbiji i jedan od najvećih administrativnih, edukativnih i industrijskih centara koji broji preko 150.000 stanovnika. Posmatrajući pomenutu lokaciju, glavni izvori zagađenja su:

- Toplana koja je pretežno koristila ugalj kao energent (od grejne sezone 2022/2023 toplana je u potpunosti prešla na prirodni gas)
- Individualna ložišta u kombinaciji sa zgradama niske energetske efikasnosti;
- Intezivan saobraćaj karakterističan po nedostatku kružnih tokova;
- Izvori zagađenja van grada, kao što je obližnji auto-put A1 (Beograd - Niš).

Na slici 1 je prikazana varijacija koncentracije  $PM_{2,5}$  u gradu Kragujevcu za period od 1.03.2021. do 28.02.2022. godine.

$PM_{2,5}$  koncentracije su merene na 3 različite lokacije u Kragujevcu koristeći senzor „SDS 011“ postavljen od strane građana, čija je verodostojnost proverena u radu [17]. Ovaj rad je pokazao da ovaj senzor ima veliki potencijal za stvaranje niskobudžetne mreže za monitoring zagađenja, jer je pokazivao visoku tačnost sve dok se ne dostigne vrednost relativne vlažnosti vazduha od 70%, nakon koje su se javljale blage greške pri merenju. Sva merenja sa pomenutog senzora su prikupljena sa sensor.community [18] otvorene internet baze podataka.



Slika 1. Varijacija koncentracije  $PM_{2,5}$  u Kragujevcu za period od 1.3.2021. do 28.2.2022.

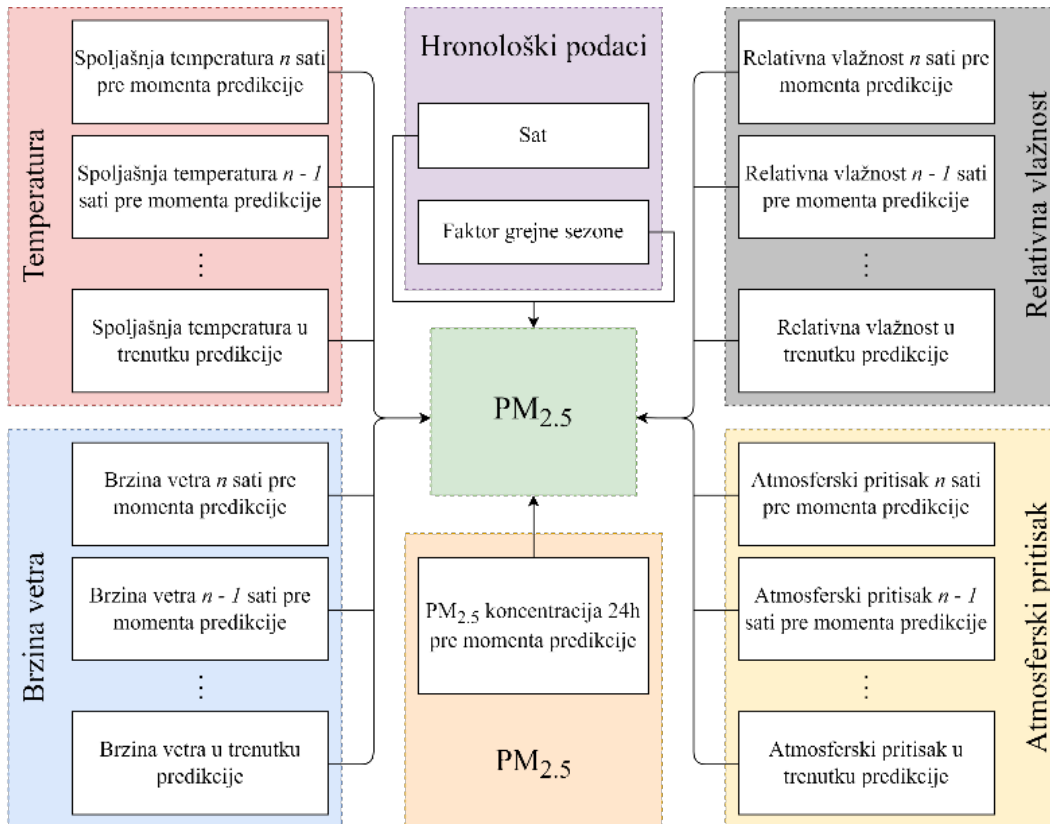


Slika 2. Metodologija istraživanja

Meteorološki podaci za Kragujevac za posmatrani period preuzeti su sa internet stranice „Nasa Data Access Viewer“ [19] čija je verodostojnost proverena u radu [20] na primeru predikcije solarne iradijacije.

Na slici 2 je prikazana metodologija istraživanja.

Prvi koraci istraživanja jesu prikupljanje podataka i njihovo pretprocesiranje. Podaci su skinuti sa archive.sensor.community [18] i Nasa Data Access Viewer [19] internet stranica, izvršeno je njihovo časovno osrednjavanje i korekcija podataka koji nedostaju, i na kraju su svi podaci spojeni u jednu bazu podataka. Baza podataka je zatim podeljena na bazu za treniranje, koja sadrži 70% nasumično odabranih podataka, i na bazu za testiranje, koja sadrži preostalih 30% podataka. Korišćeno je ukupno 4 regresionih algoritama koji su trenirani u cilju dobijanja prediktivnog modela, a to su: Random forest (RF), XGBoost (XGB), CatBoost (CB) i Light gradient boosting machine (LGBM) algoritmi. Pomenutim algoritmima je ponuđena baza za treniranje, na kojoj su učili i razvili prediktivni model. Nakon treniranja ovih regresionih algoritama, svaki od treniranih modela je testiran primenom baze za testiranje, pri čemu su njihove prediktivne performanse opisane koeficijentom determinacije ( $R^2$ ), osrednjenom apsolutnom greškom (MAE) i korenom srednje kvadratne greške (RMSE). Na osnovu pomenutih pokazatelja prediktivnih performansi modela je odabran model koji pruža najveću tačnost pri najoptimalnijem broju ulaznih varijabli.



Slika 3. Vizuelizacija pristupa problemu

Pristup samom problemu je baziran na pretpostavci da većina uticajnih parametara ima, pored trenutnog, i retrospektivni uticaj na koncentraciju  $PM_{2.5}$ . Pod retrospektivnim uticajem se podrazumeva da varijacije određenog uticajnog parametra određen broj sati pre momenta predikcije ima direktan uticaj na posmatrani trenutak u sadašnjosti. Varijabla koja opisuje vremenski dijapazon u prošlosti u kojem se uzimaju časovne varijacije svih uticajnih parametara se naziva uticajni period. Uticajni periodi ( $n$ ) koji su uzeti u razmatranje su: 48h, 24h, 12h, 8h, 4h, 2h, 1h i 0h. Vizuelizacija pristupa problemu je data na slici 3.

### III REZULTATI I DISKUSIJA

Za svaki od primenjenih algoritama razvijen je model koji koristi različit uticajni period parametara. U tabeli 1 su prikazani pokazatelji prediktivnih performansi za svaki algoritam pri svakom od analiziranih uticajnih perioda.

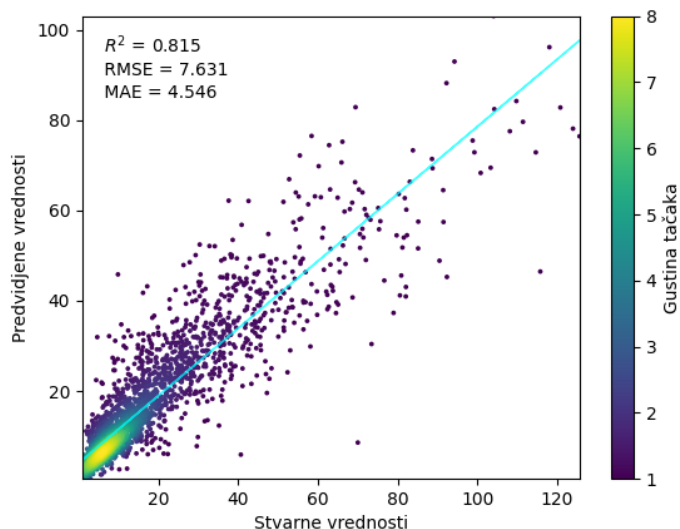
**Tabela 1.** Pokazatelji prediktivnih performansi istreniranih modela

Uticajni period		48h	24h	12h	8h	4h	2h	1h	0h
RF	$R^2$	0,78	0,76	0,75	0,74	0,72	0,71	0,70	0,71
	MAE	4,90	5,15	5,21	5,34	5,54	5,58	5,60	5,61
	RMSE	8,27	8,73	8,82	9,02	9,29	9,50	9,58	9,37
XGB	$R^2$	0,80	0,77	0,75	0,74	0,71	0,70	0,68	0,66
	MAE	4,79	5,00	5,37	5,60	5,89	5,99	6,15	6,34
	RMSE	7,94	8,51	8,90	9,07	9,57	9,72	10,10	10,40
CB	$R^2$	0,83	0,81	0,78	0,75	0,72	0,70	0,69	0,69
	MAE	4,45	4,74	5,16	5,44	5,78	5,93	6,04	6,04
	RMSE	7,36	7,71	8,38	8,80	9,41	9,71	9,90	9,84
LGBM	$R^2$	0,80	0,78	0,75	0,74	0,70	0,69	0,67	0,68
	MAE	4,93	5,10	5,44	5,52	5,98	5,97	6,08	6,12
	RMSE	7,98	8,32	8,82	9,04	9,76	9,88	10,12	10,07
Broj ulaznih		199	103	55	39	23	15	11	7

Na osnovu tabele 1 je uočljivo da su prediktivne performanse svih korišćenih algoritama slične, pri čemu je najbolje rezultate dao CB. Takođe je uočljivo da se sa povećanjem uticajnog perioda, samom algoritmu nudi veći broj ulaznih varijabli za učenje, pa je samim tim njegova tačnost veća. Preporučeni model jeste CatBoost algoritam u kombinaciji sa uticajnim periodom od 8h koji je dao relativno visok  $R^2$  faktor (0.75) pri relativno niskom broju ulaznih varijabli (39). U cilju poboljšanja prediktivnih performansi preporučenog modela potrebno je optimizovati hiperparametre CB regresora. Prema zvaničnoj internet stranici [catboost.ai](https://catboost.ai) [19], tri hiperparametra ovog algoritma imaju značajan uticaj na njegove performanse, a to su: dubina stabla (Tree depth), broj iteracija (Number of iterations) i stopa učenja (Learning rate). Što se tiče dubine stabla, sa povećanjem dubine bi trebalo da se poveća tačnost predikcije, pri čemu su preporučene vrednosti ovog hiperparametra od 4 do 10, zbog čega se usvaja da je dubina stabla 10. Hiperparametri stepen učenja i broj iteracija su zavisni jedan od drugog, i preporučuje se da se pri manjem stepenu učenja smanji broj iteracija i obratno. Kako je predodređena vrednost stepena učenja samog algoritma 0.03, usvaja se da je broj iteracija 5000.

Optimizovani model baziran na CatBoost algoritmu i uticajnom periodu od 8h je postigao značajan skok u svojim prediktivnim performansama. On je rezultovao višim koeficijentom determinacije (0,815), nižom osrednjenom apsolutnom greškom

(4.55) i nižim korenom srednje kvadratne greške (7.63). Na slici 4 je prikazan scatter dijagram koji pokazuje odnos stvarnih i predviđenih vrednosti za preporučeni model.



**Slika 4.** Odnos stvarnih i predviđenih vrednosti za preporučeni model

Na osnovu prikaza sa slike 4 je moguće zaključiti sledeće:

- najveća gustina tačaka je oko linije trenda, što ukazuje na relativno dobre prediktivne performanse preporučenog modela,
- postoji relativno mali udeo raspršenih tačaka koji ukazuje na moguće, ali ne toliko česte, greške prediktivnog modela.

U tabeli 3 je izvršeno poređenje pokazatelja prediktivnih performansi preporučenog modela i modela razvijenih u drugim istraživanjima koji su se bavili sličnim problemom.

**Tabela 2.** Komparacija prediktivnih performansi preporučenog modela i modela razvijenih u drugim istraživanjima

Literatura	Lokacija	Algoritam	$R^2$	MAE	RMSE
[22]	Tjencin	XGB	0,95	11,77	16,29
[14]	Delhi	XGB	0,92	25,84	15,02
<b>Ovaj rad</b>	<b>Srbija</b>	<b>CB</b>	<b>0,815</b>	<b>4,55</b>	<b>7,63</b>
[16]	Kina	LGBM	0,8	/	19,8
[23]	Kina	RF	0,78	28	/
[24]	Indija	RF	0,69	30,78	44,69

Na osnovu rezultata datih u tabeli 3 moguće je zaključiti:

- posmatrajući koeficijent determinacije, preporučeni model prevazilazi većinu modela razvijenih u pregledanim radovima;
- posmatrajući pokazatelje greški (MAE i RMSE), preporučeni model prevazilazi sve modele razvijene u pregledanim radovima, što može biti izazvano ili boljim prediktivnim performansama modela, ili manjim vrednostima koncentracija  $PM_{2.5}$  na posmatranoj lokaciji;
- uzimajući u obzir da je preporučeni model učen na podacima izmerenim niskobudžetnim senzorima postavljenih od strane građana, dok su modeli u

pregledanim radovima učeni na podacima izmerenim visokobudžetnim sensorima postavljenih od strane države, preporučeni model je pokazao veoma dobre prediktivne performanse.

Na osnovu priloženih rezultata može se zaključiti da je prikazanom metodologijom razvijen model zadovoljavajućih prediktivnih performansi koji se može koristiti u svrhe blagovremenog informisanja građana o budućim koncentracijama PM<sub>2.5</sub> u vazduhu.

#### IV ZAKLJUČAK

Sa ciljem da se ublaže posledice koje PM<sub>2.5</sub> imaju na zdravlje ljudi, ovaj rad je analizirao potencijal algoritama mašinskog učenja za predikciju časovnih koncentracija PM<sub>2.5</sub> u vazduhu. Prediktivni modeli su učeni na podacima koji su izmereni niskobudžetnim laserskim sensorima za merenje PM<sub>2.5</sub> u vazduhu („SDS 011“) postavljenih od strane građana a koji su skinuti sa sensor.community otvorene baze podataka i meteorološkim podacima koji su skinuti sa „Nasa Data Access Viewer“ internet stranice. Skinuti podaci su prikupljeni u jednu bazu podataka, osrednjeni su na časovnom nivou i izvršeno je korigovanje podataka koji nedostaju.

Analizirane su prediktivne performanse četiri algoritma mašinskog učenja: Random forest, XGBoost, CatBoost i Light gradient boosting machine. Odabir pomenutih algoritama je izvršen na primeru drugih radova koji su se bavili sličnim problemom. Pristup problemu je baziran na činjenici da varijacija svih meteoroloških parametara pre momenta predikcije takođe ima uticaj na koncentraciju PM<sub>2.5</sub> u momentu predikcije. Vodeći se ovom pretpostavkom, uveden je parametar koji definiše vremenski interval u prošlosti (u odnosu na trenutak predikcije) u kome se uzimaju časovne vrednosti svakog od meteoroloških parametara koji se naziva uticajni period. Analizirani uticajni periodi su 48h, 24h, 12h, 8h, 4h, 2h, 1h i 0h. Analizom prediktivnih performansi svakog od pomenutih algoritama, pri svakom od pomenutih uticajnih perioda, zaključeno je da CB daje najbolje performanse. Uzimajući u obzir vrednost koeficijenta determinacije i broja ulaznih varijabli, preporučuje se model baziran na CB algoritmu i uticajnom periodu od 8h koji je okarakterisan sa 39 ulaznih varijabli. Sa ciljem poboljšanja prediktivnih performansi preporučenog modela, izvršena je optimizacija njegovih hiperparametara. Optimizovani preporučeni model je dostigao relativno visoku vrednost koeficijenta determinacije (0,815) i relativno niske vrednosti osrednjene prosečne greške (4,55) i korena srednje kvadratne greške (7,63).

Poređenjem preporučenog modela, koji je baziran na niskobudžetnim sensorima, sa modelima razvijenim u drugim istraživanjima, koji su bazirani na visokobudžetnim sensorima, zaključuje se da dati model daje veoma dobre prediktivne performanse i da se može primeniti za blagovremenu predikciju koncentracija PM<sub>2.5</sub>.

Predmetno istraživanje pruža univerzalnu metodologiju za precizno predviđanje koncentracije PM<sub>2.5</sub> koja je primenjena na primeru grada Kragujevca za period od 1. 3. 2021. do 28. 2. 2022. godine. Usled različitosti klimatskih uslova, ljudskih

aktivnosti, saobraćaja i industrije, svaka lokacija predstavlja jedinstven slučaj koji bi zahtevao zasebno učenje modela. Analogno tome, svaka promena ovih uslova za neku lokaciju bi zahtevala ponovno učenje modela na novim podacima koji uzimaju u obzir predmetnu promenu.

#### LITERATURA/REFERENCES

- [1] Jurišević, N., Stojadinović, M., Končalović, D., Josijević, M., Gordić, D. Students' Perceptions of Air Quality - an Opportunity for More Sustainable Urban Transport in the Medium-sized University city in the Balkans, *Tehnika*, Vol. 78, No. 4, pp. 455-463, 2023. <https://doi.org/10.5937/tehnika2304455J>
- [2] World Health Organization, Air quality and health, <https://www.who.int/teams/environment-climate-change-and-health/air-quality-and-health/health-impacts/types-of-pollutants> [pristupljeno 30.4.2023]
- [3] Nansai, K., Tohno, S., Chatani, S., Kanemoto, K., Kagawa, S., Kondo, Y., Takayanagi, W., Lenzen, M. Consumption in the G20 nations causes particulate air pollution resulting in two million premature deaths annually. *Nature Communications*, Vol. 12, No. 6286, 2021. <https://doi.org/10.1038/s41467-021-26348-y>
- [4] United States Environmental Protection Agency. Particulate Matter (PM) Basics, <https://www.epa.gov/pm-pollution/particulate-matter-pm-basics> [pristupljeno 30.4.2023]
- [5] World Health Organization, Regional Office for Europe. Health effects of particulate matter: policy implications for countries in eastern Europe, Caucasus and central Asia. 2013. <https://iris.who.int/handle/10665/344854> [pristupljeno 30.4.2023]
- [6] Đurišić, Ž., Škrbić, B., Potencijal energije sunca i vetra za strateško planiranje dekarbonizacije proizvodnje električne energije u Srbiji, *Energija, ekonomija, ekologija*, Vol. 24, No. 4, pp. 1-11, 2022. <https://doi.org/10.46793/EEE22-4.01D>
- [7] Chen, Z., Chen, D., Zhao, C., Kwan, M., Cai, J., Zhuang, Y., Zhao, B., Wang, X., Chen, B., Yang, J., Li, R., He, B., Gao, B., Wang, K., Xu, B. Influence of meteorological conditions on PM<sub>2.5</sub> concentrations across China: A review of methodology and mechanism, *Environment International*, Vol. 139, No. 105558, 2020. <https://doi.org/10.1016/j.envint.2020.105558>
- [8] Megaritis, A.G., Fountoukis, C., Charalampidis, P.E., Denier Van Der Gon, H.A.C., Pilinis, C., Pandis, S.N. Linking climate and air quality over Europe: Effects of meteorology on PM<sub>2.5</sub> concentrations. *Atmospheric Chemistry and Physics*, Vol. 14, No. 18, pp. 10283-10298, 2014. <https://doi.org/10.5194/acp-14-10283-2014>
- [9] Zalakevičiute, R., López-Villada, J., Rybarczyk, Y. Contrasted effects of relative humidity and precipitation on urban PM<sub>2.5</sub> pollution in high elevation urban areas, *Sustainability*, Vol. 10, No. 6, pp. 2064, 2018. <https://doi.org/10.3390/su10062064>
- [10] Wang, J., Ogawa, S. Effects of meteorological conditions on PM<sub>2.5</sub> concentrations in Nagasaki, Japan, *International Journal of Environmental Research and Public Health*, Vol. 12, No. 8, pp. 9089-9101, 2015. <https://doi.org/10.3390/ijerph120809089>
- [11] Hernandez, G., Berry, T.-A., Wallis, S.L., Poyner, D. Temperature and humidity effects on particulate matter concentrations in a sub-tropical climate during winter, in *Proc. International Proceedings of Chemical, Biological and Environmental Engineering, (ICECB 2017)*, 2017. <https://doi.org/10.7763/IPCBE.2017.V102.10>
- [12] Chang, L.T.C., Scorgie, Y., Duc, H.N., Monk, K., Fuchs, D., Trieu, T. Major source contributions to ambient PM<sub>2.5</sub> and exposures within the New South Wales Greater Metropolitan Region, *Atmosphere*, Vol. 10, No.3, pp. 138, 2019. <https://doi.org/10.3390/atmos10030138>
- [13] Doreswamy, H.K.S., Km, Y., Gad, I. Forecasting air pollution particulate matter (PM<sub>2.5</sub>) using machine learning regression models, *Procedia Computer Science*, Vol. 171, pp. 2057-2066, 2020. <https://doi.org/10.1016/j.procs.2020.04.221>
- [14] Kumar, S., Mishra, S., Singh, S.K. A machine learning-based model to estimate PM<sub>2.5</sub> concentration levels in Delhi's atmosphere, *Heliyon*, Vol. 6, No. 11, pp. E05618, 2020. <https://doi.org/10.1016/j.heliyon.2020.e05618>
- [15] Shahriar, S.A., Kayes, I., Hasan, K., Hasan, M., Islam, R., Awang, N.R., Hamzah, Z., Rah, A., Salam, M. Potential of ARIMA-ANN, ARIMA-SVM,

- DT and CatBoost for Atmospheric PM<sub>2.5</sub> forecasting in Bangladesh, *Atmosphere*, Vol. 12, No. 1, pp. 100, 2021. <https://doi.org/10.3390/atmos12010100>
- [16] Zhong, J., Zhang, X., Gui, K., Wang, Y., Che, H., Shen, X., Zhang, L., Zhang, Y., Sun, J., Zhang, W. Robust prediction of hourly PM<sub>2.5</sub> from meteorological data using LightGBM, *National Science Review*, Vol. 8, No. 10, nwaa307, 2021. <https://doi.org/10.1093/nsr/nwaa307>
- [17] Liu, H.Y., Schneider, P., Haugen, R., Vogt, M. Performance assessment of a low-cost PM 2.5 sensor for a near four-month period in Oslo, Norway, *Atmosphere*, Vol. 10, No. 2, pp. 41, 2019. <https://doi.org/10.3390/atmos10020041>
- [18] Archive - Sensor.Community. <https://archive.sensor.community/> [pristupljeno 03.05.2023]
- [19] Data Access Viewer. <https://power.larc.nasa.gov/data-access-viewer/> [pristupljeno 03.05.2023]
- [20] Shuvho, M.B.A., Chowdhury, M.A., Ahmed, S., Kashem, M.A. Prediction of solar irradiation and performance evaluation of grid connected solar 80KWp PV plant in Bangladesh, *Energy Reports*, Vol. 5, pp. 714-722, 2019. <https://doi.org/10.1016/j.egy.2019.06.011>
- [21] CatBoost. Parameter tuning, <https://catboost.ai/en/docs/concepts/parameter-tuning> [pristupljeno 10.4.2023]
- [22] Pan, B. Application of XGBoost algorithm in hourly PM<sub>2.5</sub> concentration prediction, in *Proc. IOP Conference Series: Earth and Environmental Science*, Vol. 113, *3rd International Conference on Advances in Energy Resources and Environment Engineering 8-10 December 2017, Harbin, China, 2018*. <https://doi.org/10.1088/1755-1315/113/1/012127>
- [23] AlDaweesh, S.A. Predicting hourly particulate matter (PM<sub>2.5</sub>) concentrations using meteorological data, in *Proc. International Conference of Computing, Electronics & Communications Engineering (ICCECE)*, pp. 136-140, London, 2019. <https://doi.org/10.1109/ICCECE46942.2019.8941696>
- [24] Sihag, P., Kumar, V., Afghan, F.R., Pandhiani, S.M., Keshavarzi, A. Predictive modeling of PM<sub>2.5</sub> using soft computing techniques: case study-Faridabad, Haryana, India, *Air Quality, Atmosphere & Health*, Vol. 12, pp. 1511-1520, 2019. <https://doi.org/10.1007/s11869-019-00755-z>

## AUTORI/AUTHORS

**Filip Nastić** - istraživač-pripravnik, Fakultet inženjerskih nauka Univerziteta u Kragujevcu, [filip.nastic@uni.kg.ac.rs](mailto:filip.nastic@uni.kg.ac.rs), ORCID [0000-0002-2164-6658](https://orcid.org/0000-0002-2164-6658)